

Binomial Probability Calculations

Dr Richard Kenderdine

Kenderdine Maths Tutoring

Calculations for binomial probability questions can be done four ways:

- (1) Exactly using the binomial distribution for the number of successful outcomes
- (2) Approximately using the Normal Distribution approximation for outcomes without a continuity correction
- (3) Approximately using the Normal Distribution approximation for outcomes with a continuity correction
- (4) Approximately using the Normal Distribution approximation for the proportion

This note will show all methods for two examples.

Example 1

The long-term average of the proportion of students at a certain school who play a musical instrument is 20%. A random sample of 80 students at the school is to be taken to determine the proportion who play a musical instrument.

- (i) *Find the expected value and standard deviation of the distribution of the sample proportion*
- (ii) *Use the Standard Normal Distribution to estimate the probability that the sample of 80 students contains at least 14 and at most 21 students who play a musical instrument.*

The question is phrased so that the appropriate solution method is (4). This solution is shown first, then the others for comparison.

(i) The expected value is 0.2 and the standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2 \times 0.8}{80}} = 0.04472$

(ii) The relevant proportions are $14/80 = 0.175$ and $21/80 = 0.2625$

$$\text{Hence } P(14 \leq X \leq 21) = P\left(\frac{0.175-0.2}{0.04472} \leq z \leq \frac{0.2625-0.2}{0.04472}\right) = P(-0.56 \leq z \leq 1.40) = 0.6315$$

Now look at the Normal approximation for outcomes without continuity correction. The expected value is $0.2(80) = 16$ and the standard deviation $\sqrt{80 \times 0.2 \times 0.8} = 3.578$

$$\text{Hence } P(14 \leq X \leq 21) = P\left(\frac{14-16}{3.578} \leq z \leq \frac{21-16}{3.578}\right) = P(-0.56 \leq z \leq 1.40) = 0.6315$$

The same result is obtained from using outcomes or proportions. This will always be true because every term in the z-transformation is divided by the sample size.

The only difference when applying the continuity correction is to use a lower limit of 13.5 and an upper limit of 21.5 when transforming to z:

$$\text{Hence } P(14 \leq X \leq 21) = P\left(\frac{13.5-16}{3.578} \leq z \leq \frac{21.5-16}{3.578}\right) = P(-0.70 \leq z \leq 1.54) = 0.6963$$

The exact result, obtained using the binomial distribution for outcomes, is a probability of 0.687, showing that the Normal approximation using the continuity correction (0.6963) provides a more accurate result. Omitting the continuity correction or, equivalently, using the z-test for proportions, estimated the probability as 0.6315, an error of about 9%. Figure 1 shows the output using GeoGebra.

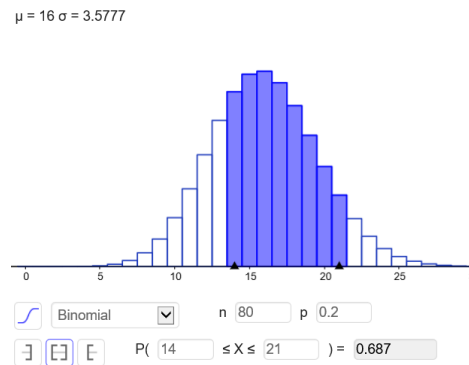


Figure 1: Binomial probability for outcomes from 14 to 21 inclusive with sample size 80 and $p = 0.2$

Increasing the sample size will give better agreement between the Normal approximation and the exact Binomial probability. For example, increase the sample size to 8000 and keep the z-scores for the proportion z-test as -0.56 and 1.40. The new standard deviation is $\sqrt{\frac{0.2 \times 0.8}{8000}} = 0.00447$, thus giving the test proportions as solutions to the equations

$$\frac{p-0.2}{0.00447} = -0.56 \quad \text{and} \quad \frac{p-0.2}{0.00447} = 1.40$$

These solutions are 0.1975 and 0.2063, representing outcomes $0.1975(8000) = 1580$ and $0.2063(8000) = 1650$

Using the Binomial distribution, the probability of obtaining outcomes from 1580 to 1650 inclusive is 0.6366, the error using the z-test for proportions now being only 0.8%. (GeoGebra couldn't handle such a large sample size, calculations were therefore obtained using *Mathematica*).

Example 2

According to Mars, 24% of all M&M plain candies are blue. Assuming that this is correct, find the probability that at least 21 out of 80 randomly selected M&Ms are blue.

First, use either the z-test for proportions or the Normal approximation without continuity correction. The sample proportion is $21/80 = 0.2625$. The standard deviations are

$$\sqrt{\frac{(0.24)(0.76)}{80}} = 0.0477 \quad \text{for proportions and}$$

$$\sqrt{80(0.24)(0.76)} = 3.8199 \quad \text{for outcomes}$$

The z-score is then $\frac{0.2625-0.24}{0.0477} = 0.471$ for proportions, or $\frac{21-0.24(80)}{3.8199} = 0.471$ for outcomes.

The probability of obtaining a z-score greater than 0.471 is 0.3188.

When using the continuity correction the z-score is $\frac{20.5-0.24(80)}{3.8199} = 0.340$ and the probability of obtaining a z-score greater than 0.340 is 0.3669.

The exact probability, using the Binomial Distribution, of obtaining 21 or more with sample size 80 and $p = 0.24$ is 0.3594.

Hence the z-test without continuity correction yields a probability (0.3188) that is more than 11% less than the true value (0.3594) while the probability obtained using the correction (0.3669) is 2% greater. These errors will reduce with a larger sample size.

Conclusion

Calculating the exact binomial probability can be tedious if there are a large number of outcomes to be included. A z-test for the proportion, or the equivalent z-test for outcomes, gives an approximate probability based on the Normal Distribution. The error can be relatively large when the sample size is small and it is better to always use the z-test with continuity correction.